

Faisabilité d'un capteur 3D basé sur l'analyse du flou optique sur les images

Thierry Simon⁽¹⁾, Christophe Simon⁽²⁾, Frédérique Bicking⁽²⁾

⁽¹⁾ Equipe PERCEVAL, IUT de Figeac, Dpt GMP, Avenue de Nayrac, 46100 Figeac

⁽²⁾ ESSTIN, Université Henri Poincaré Nancy 1, 2 Rue Jean Lamour, 54519 Vandœuvre-lès-Nancy Cedex

Thierry.Simon@univ-tlse2.fr, {Christophe.Simon;Frederique.Bicking}@esstin.uhp-nancy.fr

I. INTRODUCTION

Le développement de capteurs et notamment de capteurs intelligents pour la mesure est un axe d'initiative de recherche pertinent tant pour les laboratoires que pour l'industrie. En outre, le domaine de la mesure tridimensionnelle sans contact est particulièrement ouvert puisque les dispositifs de mesure actuels sont de type ponctuel ou linéaire et que très peu de dispositifs s'attèlent au problème du mesurage 3D direct. Dans ce domaine, les caméras sont les corps d'épreuve privilégiés, il reste alors à les mettre en œuvre habilement pour aboutir à un système de mesure en trois dimensions. De nombreuses méthodes ont été développées pour obtenir ces coordonnées. Elles reposent toutes sur une exploitation des variations des conditions de prise de vue de plusieurs images. Les paramètres du système d'acquisition ou de l'environnement lumineux, contrôlé ou non, sont les informations essentielles à l'établissement de la relation entre l'image et la scène réelle. Deux grandes catégories d'approches peuvent être définies. Les approches multi-oculaires où une même scène est captée selon plusieurs points de vue et les approches monoculaires où la scène est captée selon le même point de vue. La première catégorie est extrêmement développée mais les techniques présentent l'inconvénient majeur d'un lourd traitement de l'information pour apparier des points afin de trianguler leur position. Néanmoins, un bon appariement conduira à une mesure 3D précise. La seconde catégorie regroupe les techniques utilisant le flou optique comme information de profondeur. Le problème d'appariement n'existe plus, mais le niveau de précision dépend fortement des

conditions de prise de vue. Ces techniques se composent de deux groupes. Le premier groupe recherche dans l'image les zones de netteté et les lie à la profondeur par les relations classiques de l'optique géométrique. Nommées Depth From Focus ou DFF ([1]; [2]; [3]), elles présentent l'inconvénient majeur d'un temps prohibitif d'obtention d'une carte des profondeurs. Le second groupe exploite le flou optique perceptible sur les zones non homogènes de l'image correspondant à des contours ou à des textures. Ces techniques, appelées Depth From Defocus ou DFD ([4]; [5]; [6]), utilisent au moins deux images d'une même scène acquises avec des réglages différents des paramètres du système optique. Les principales techniques exploitent soit le contenu spatial de l'image par les caractéristiques géométriques ([4]) ou la forme des objets dans la scène ([2]) soit le contenu fréquentiel par une analyse de Fourier ([7]). Cette analyse est toujours très coûteuse en temps de calcul. Elle est donc peu appropriée au développement d'un capteur. C'est pourquoi nous avons préféré une approche spatiale localisée sur des éléments caractéristiques des objets de la scène observée. Dans cet article, nous présentons le développement théorique de notre méthode de perception de la profondeur dans une scène en rappelant le processus de formation des images floues, en explicitant en quel sens le flou optique est une information sur la profondeur et en précisant les modèles mathématiques utilisés. Nous analysons ensuite les constituants de la méthode, notamment en terme de sélection des opérateurs de traitement de l'image et du modèle de la réponse impulsionnelle du système optique. Nous discutons dans une quatrième partie de l'influence

de ces modèles et des opérateurs utilisés sur la précision des résultats de mesure de profondeurs et sur les temps de calcul. Enfin, nous évaluons l'opportunité de mettre en oeuvre cette approche sur une architecture matérielle pour le développement d'un capteur monoculaire permettant la perception de la profondeur, appelé le capteur 3D.

II. DÉVELOPPEMENT THÉORIQUE DE LA MÉTHODE DE PERCEPTION DE LA PROFONDEUR

Les techniques DFD permettent la détermination de la profondeur d'un objet en utilisant au moins deux images acquises avec un seul système optique dans des conditions de réglage différentes. Si l'image d'une scène est acquise par une lentille réelle, les points de la surface de la scène situés à une distance particulière de la lentille seront focalisés, ils apparaîtront nets sur l'image. Les points à d'autres distances seront défocalisés, flous au sens optique. Leur degré de flou dépend de la distance les séparant de la caméra. C'est sur ce principe que repose nos développements.

A. Le flou optique porteur de l'information de profondeur

La méthode DFD exploite l'effet physique produit sur les propriétés d'images par la modification des paramètres d'acquisition tels que la longueur focale, la mise au point et l'ouverture du diaphragme. On cherche à établir une ou plusieurs relations entre la profondeur, les paramètres de la caméra et la quantité de flou optique dans les images.

Pour des raisons de simplification, nous assimilons l'objectif de la caméra à une lentille convergente mince et de longueur focale équivalente.

Pour une caméra avec une lentille de longueur focale f , la relation entre la position d'un point dans la scène et la position de son image focalisée (ou nette) est donnée par la loi de conjugaison :

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f} \quad (1)$$

où s_o est la distance entre le plan principal (plan lentille) et l'objet et s_i est la distance entre le plan principal et le plan image.

La figure 1 illustre le processus géométrique de formation des images. Tous les rayons lumineux

émis par un objet O et interceptés par la lentille sont réfractés pour converger en un point sur le plan image. Chaque point dans la scène est projeté en son point conjugué sur le plan focal formant sur celui-ci une image focalisée ou nette. Si le plan capteur ne coïncide pas avec le plan focal d'une distance δ , l'image formée sera une tache appelée "cercle de confusion". Ainsi, le point objet devient un disque avec un diamètre valant $2R$ si l'ouverture du diaphragme est circulaire.

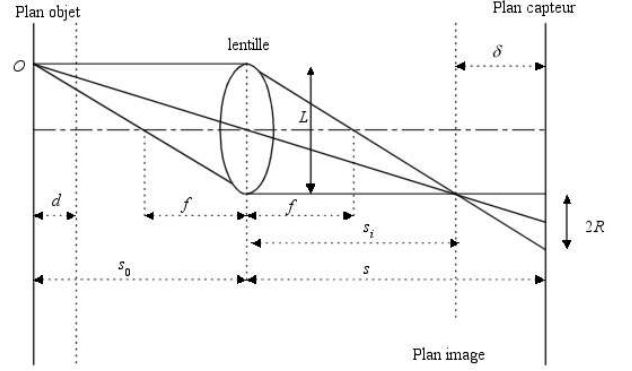


Figure 1 – Principe de formation des images selon l'optique géométrique.

Le diamètre du cercle répond à l'équation 2 :

$$2R = \frac{Ls}{\left(\frac{1}{f} - \frac{1}{s_o} - \frac{1}{s_i}\right)} \quad (2)$$

Pour $\delta \geq 0$ (fig. 1), le plan capteur de la caméra est derrière le plan focal et pour $\delta < 0$, il est devant. La figure 2 montre l'évolution du diamètre du cercle de confusion en fonction de la distance du plan objet s_o pour les réglages de paramètres suivants $s_i = 70 \text{ mm}$, $f = 60 \text{ mm}$ et pour différentes ouvertures du diaphragme L .

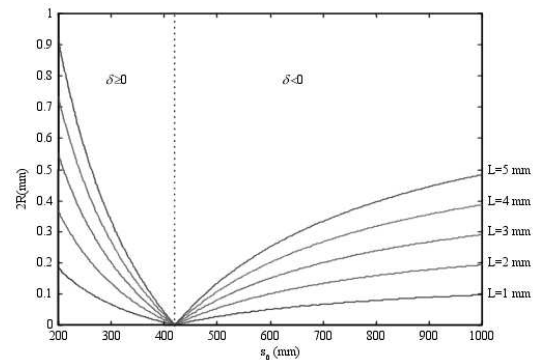


Figure 2 – Variation du diamètre du cercle flou en fonction de s_o .

La variation de signe de δ peut conduire à une estimation non unique du flou optique donc de

la profondeur puisque à une valeur de $2R$ peut correspondre deux valeurs de s_o . Pour lever cette ambiguïté, il suffit de placer le plan focal soit en avant plan soit en arrière plan de la scène. Ce choix sera guidé par un compromis à faire entre l'étendue de profondeur et la sensibilité du "capteur tridimensionnel".

A partir des relations d'optique, la distance à un objet ou profondeur dans une scène peut s'exprimer par l'équation (3) :

$$\begin{aligned} s_o &= \frac{fs}{s-f-2R\frac{f}{L}} \text{ pour } \delta < 0 \\ s_o &= \frac{fs}{s-f+2R\frac{f}{L}} \text{ pour } \delta \geq 0 \end{aligned} \quad (3)$$

Cette expression est à la base de toutes les méthodes qui utilisent le flou optique comme information de profondeur. Le lien entre l'information floue caractérisée par l'étendue du cercle de confusion et l'information de profondeur exprimant la distance objet - lentille est alors immédiat pour une profondeur s_o avec la relation (4).

$$\frac{1}{s_o} = \frac{k(2R)}{m} - \frac{c}{m} \quad (4)$$

où k , c et m des constantes représentant des réglages spécifiques de la caméra. Elle sont déterminées par une procédure appropriée de calibrage.

En général, on ne mesure pas directement la taille de la tache car celle-ci n'est pas un disque parfait. La distribution de son intensité lumineuse est intimement liée au comportement ondulatoire du système optique. Le modèle utilisé pour représenter cette tache sera caractéristique de la fonction de transfert du système optique (FTO) qui permet d'établir un lien entre R et l'étalement de la tache image. La FTO, caractéristique d'une profondeur permet d'exprimer la relation entre un plan focalisé et non focalisé par $I_F(u, v) = I_N(u, v) * FTO(u, v)$ où $I_F(u, v)$ est la transformée de Fourier de l'image floue et $I_N(u, v)$ celle de l'image focalisée. Cette relation s'écrit dans le domaine spatiale $i_F(i, j) = i_N(i, j) \otimes_{2D} h(i, j)$ où \otimes_{2D} est l'opérateur de convolution bidimensionnel et $h(i, j)$ la réponse impulsionnelle du système optique obtenue par la transformée de Fourier inverse de la FTO.

La connaissance de la réponse impulsionnelle du système optique appelée PSF (Point Spread

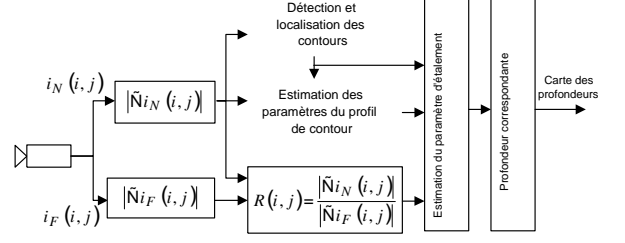


Figure 3 – Schéma de principe

Function ou fonction d'étalement du point) permet d'obtenir le paramètre d'étalement Pe_{s_o} pour différentes profondeurs s_o . On trouve la relation (5) où les constantes m et c sont obtenues par calibrage.

$$\frac{1}{s_o} = \frac{Pe_{s_o}}{m} - \frac{c}{m} \quad (5)$$

B. Principe de la méthode

L'information de profondeur existe à travers le flou optique. Il est alors nécessaire d'identifier certaines caractéristiques de l'image pour lesquelles les effets du flou sont mesurables comme par exemple des discontinuités d'intensité. Ces discontinuités spatiales représentent parfois les contours d'objets observés. Afin d'appliquer la méthode, il est nécessaire d'identifier ces contours et de les positionner.

Le principe général de la méthode, présenté sur la figure 3) est similaire à celui proposé par Pentland ([4]; [8]). On retient l'acquisition d'une image nette en tout point avec un iris fermé qui fournit facilement la position des contours grâce à un opérateur de gradient. En faisant l'acquisition d'une seconde image floue (iris ouvert), le contenu informationnel de la scène est isolé. Une estimation du flou est obtenue à partir du module des gradients de luminance. En établissant le rapport des modules des gradients des images nette et floue, qui permet de donner la même importance à chaque contour, il est possible d'identifier le paramètre de la PSF.

Nous lions ensuite l'estimation de ce paramètre à la profondeur en utilisant la relation (5). Si le paramètre d'étalement est déterminé en chaque pixel appartenant aux contours des objets de la scène, nous obtenons une carte des profondeurs.

Le rapport des modules des gradients des deux images s'écrit :

$$R(i, j) = \frac{|\nabla i_N(i, j)|}{|\nabla i_F(i, j)|} = \frac{|\nabla i_N(i, j)|}{|\nabla(i_N(i, j) \otimes h(i, j))|} \quad (6)$$

Où ∇ représente l'opérateur gradient.

L'originalité de l'approche est de considérer la discontinuité spatiale sous une forme monodimensionnelle plutôt que bidimensionnelle. On utilise un modèle type de contours parmi les plus courants ([9]) que l'on localise dans l'image nette.

Considérons un contour $c_n(x)$ pris dans la direction de l'axe x perpendiculaire au contour de l'objet. On le définit par l'expression (7) comme une rampe d'amplitude $b - a$, de longueur ε .

$$c_n(x) = \begin{cases} a & x < x_0 \\ a + \left(\frac{b-a}{\varepsilon}\right)(x - x_0) & x_0 \leq x \leq x_0 + \varepsilon \\ b & x > x_0 + \varepsilon \end{cases} \quad (7)$$

Le gradient du profil du contour dans l'image floue est donné par :

$$\nabla c_f(x) = \nabla c_n(x) \otimes h(x) \quad (8)$$

si la PSF $h(i, j)$ est de symétrie circulaire.

Le rapport des modules des gradients devient alors :

$$R(x) = \frac{|\nabla c_n(x)|}{|\nabla c_f(x)|} = \frac{|\nabla c_n(x)|}{|\nabla c_n(x) \otimes h(x)|} \quad (9)$$

En traitement numérique des images de nombreuses approximations de l'opérateur gradient permettent d'obtenir une forme analytique de l'équation (9) plus ou moins simple. Nous avons choisi l'opérateur de Prewitt pour sa simplicité d'application en contexte temps réel et sa relative tenue au bruit puisqu'il intègre un filtrage par moyennage ([10]). On peut alors écrire la relation :

$$R(x) = \begin{cases} \frac{1}{h(x-x_0)+h(x-(x_0+\varepsilon))+2\sum_{u=1}^{\varepsilon-1}h(x-(x_0+u))} & \text{pour } x = x_0 \text{ et } x = x_0 + \varepsilon \\ \frac{2}{h(x-x_0)+h(x-(x_0+\varepsilon))+2\sum_{u=1}^{\varepsilon-1}h(x-(x_0+u))} & \text{pour } x \in]x_0, x_0 + \varepsilon[\end{cases}$$

Avec cette expression du rapport des modules des gradients, nous sommes capables d'estimer $(\varepsilon + 1)$ valeurs du paramètre d'étalement pour tout $x \in [x_0; x_0 + \varepsilon]$. On obtient une valeur

finale par la moyenne des valeurs estimées. La méthode s'en trouve un peu alourdie en temps de calcul mais l'estimation est globalement meilleure. Les détails théoriques sont donnés dans ([11]).

III. DISCUSSION SUR LES MODÈLES UTILISÉS

A. Opérateurs de traitement des images

Nous avons constaté que le filtrage de l'opérateur de Prewitt n'était pas assez performant. Aussi, on lui préfère l'opérateur de Sobel ([10]) pour sa meilleure tenue au bruit. L'expression thorique de rapport des modules des gradients donne en (B.) est valable pour ces deux opérateurs. En revanche, le coût de calcul est plus important que lors de l'utilisation de l'opérateur de Prewitt car on augmente le nombre d'opérations en raison des coefficients non unitaires introduits dans les masques de convolution.

L'opérateur optimal de Canny-Deriche, opérateur de référence en matière de calcul de gradient, repose sur une forme exponentielle à support non borné. Son utilisation est difficile dans notre approche locale. De plus, le coût de calcul devient conséquent et réduit l'opportunité d'une application temps réel malgré ses performances.

B. Modèles de PSF

La PSF dépend autant des propriétés des matériaux de l'optique (indice de réfraction), que de la forme géométrique de la lentille (longueur focale) et des paramètres de prise de vue (distance de l'objet, diaphragme, éclairage). Un modèle réel prenant en compte à la fois tous les aspects de l'optique géométrique, les effets de la diffraction et les défauts de lentilles n'existe pas. Nous avons précisé précédemment que l'analyse du paramètre d'étalement pouvait se faire en une dimension si la PSF est de symétrie circulaire.

Les trois principaux modèles de réponse impulsionnelle de l'optique sans aberration sont le modèle de Hopkins ([12]), le modèle gaussien et le modèle géométrique.

1. Le modèle de Hopkins

Ce modèle prend bien en compte le phénomène de diffraction. Il est défini par la relation (10) :

IV. RÉSULTATS DE MESURE

$$h(x, y, \lambda) = \left(\frac{2}{a}\right)^2 [U_1^2(a, b) + U_2^2(a, b)] I_0 \quad (10)$$

avec les fonctions de Lommel :

$$U_n(a, b) = \sum_{k=0}^{\infty} (-1)^k \left(\frac{a}{b}\right)^{n+2k} J_{n+2k}(b),$$

$I_0 = \left[\frac{\pi}{\lambda} \left(\frac{L}{2f}\right)^2\right]^2 I_e$ l'intensité au point origine avec I_e l'intensité émise et, $a = \frac{2\pi}{\lambda} \left(\frac{L}{2f}\right)^2 (s - s_i)$ et $b = \frac{2\pi}{\lambda} \left(\frac{L}{2f}\right)^2 \sqrt{x^2 + y^2}$.

$J_m(b)$ est une fonction de Bessel de première espèce d'ordre m et λ est une longueur d'onde. On peut remarquer que ce modèle dépend de la longueur d'onde λ . En lui affectant une valeur moyenne prise sur le spectre de réponse du capteur CCD, nous perdons peu sur l'effet globale du flou. Néanmoins, la complexité de cette forme analytique en fait un modèle peu utilisé.

2. Le modèle gaussien

Le modèle gaussien 2D avec un écart-type σ_{s_o} correspondant à une profondeur s_o est donné par l'équation (11) :

$$h(x, y) = \frac{1}{2\pi\sigma_{s_o}^2} e^{-\frac{x^2 + y^2}{2\sigma_{s_o}^2}} \quad (11)$$

Il est une bonne approximation du modèle réel dans sa forme spatiale mais est significativement différent au delà du premier lobe dans le domaine fréquentiel. Cependant, sa forme analytique simple en fait le modèle le plus utilisé.

3. Le modèle géométrique

Pour des raisons de simplicité, on peut choisir de ne pas tenir compte du comportement ondulatoire de l'optique. On se retrouve alors sur la base de l'optique géométrique avec le modèle "Pillbox". La PSF est alors une constante répartie sur une surface circulaire. (relation 12) :

$$h(x, y) = \begin{cases} \frac{1}{\pi R^2} & \sqrt{x^2 + y^2} \leq R \\ 0 & \text{ailleurs} \end{cases} \quad (12)$$

Le modèle simple est parfois utilisé dans les travaux d'optique, notamment lorsque la quantité de flou est grande.

Pour évaluer les performances d'estimation du paramètre d'étalement par notre méthode, nous avons synthétisé une image constituée des quatre rectangles de teintes (niveaux de gris) différentes, notés de 1 à 4 (fig. 4). Le rectangle 3 montre une occlusion qui est caractéristique d'une complexité classique du problème de perception de la profondeur. Ces rectangles ont été rendus artificiellement flous avec les trois valeurs de paramètres d'étalement suivantes : $\sigma_{s_{o1}} = 1.2$, $\sigma_{s_{o2}} = \sigma_{s_{o3}} = 1.8$ et $\sigma_{s_{o4}} = 2.6$ qui peuvent correspondre aux profondeurs

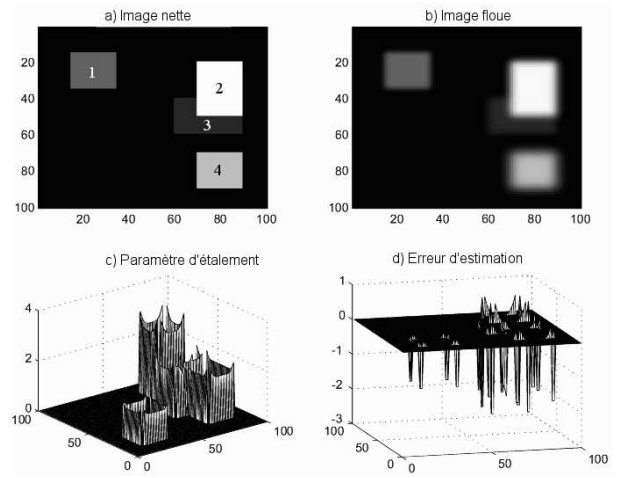


Figure 4 – Estimation locale du paramètre d'étalement

L'estimation est globalement correcte même lorsqu'il y a occlusion de formes. Cependant, aux environs des jonctions de contours, les erreurs d'estimation du paramètre d'étalement sont importantes. Cela est dû à une contribution mutuelle des flous de ces contours puisque nous ne respectons pas l'équivalence 2D vers 1D qui nous a permis d'écrire l'équation (9).

V. ARCHITECTURE MATÉRIELLE

Aspect optique : l'intérêt majeur de la stéréovision monoculaire par rapport aux solutions multioculaires est d'éliminer le problème d'appariement de points pour appliquer le principe de triangulation. Néanmoins, pour développer un capteur 3D sur le principe détaillé précédemment, il nous faut mettre en oeuvre un système optique délivrant une image nette et son équivalente floue de la même scène. Pentland ([8]) a proposé une structure distribuant l'image

de la scène sur deux CCD par deux chemins optiques différents. Il reste à intégrer ce système sur des montures standards.

Aspect électronique : le développement d'un tel capteur impose des contraintes sur le plan électronique. Les deux CCD doivent faire l'acquisition des images dans un laps de temps court si on veut éliminer les problèmes de correspondance de contours lors de mouvements d'objets. De plus, les CCD et leur électronique doivent avoir des comportements similaires par rapport au bruit afin d'introduire un minimum d'interférences. Pour pallier ces problèmes, il suffit de faire l'acquisition d'images de taille 512x512 sur un capteur de plus grande taille par exemple 2029x2044 (KODAK MEGAPLUS 4.2i). Enfin, le capteur doit être doté d'un système numérique rapide de type DSP afin de traiter simultanément les deux images.

Capteur Intelligent : l'intégration électronique d'un DSP permet d'envisager des fonctions de communication autorisant l'interaction avec le capteur. Cette fonctionnalité nous permet d'une part de paramétrer le capteur notamment pour des recalages d'images ou d'éclaircissement et d'autre part de récupérer à distance les cartes de profondeurs pour une utilisation dans des applications industrielles ou de laboratoire.

VI. CONCLUSION

Le développement théorique montre la faisabilité d'un capteur 3D monoculaire par l'exploitation du flou optique. Néanmoins, les aspects optiques et électronique du développement sont non négligeables en terme de coût et de temps. Toutefois, de nombreuses applications trouveront bénéfices à disposer d'un tel capteur. On relèvera notamment des applications de reverse-engineering 3D, tous les problèmes d'analyse et de reconstruction de scènes ou encore des applications d'assistance à la conduite de véhicules ...

REFERENCES

- [1] P. Grossman, "Depth from focus," *Pattern Recognition Letters*, vol. 5, pp. 63–69, 1987.
- [2] T. Darrell and K. Wohn, "Depth from focus using a pyramid architecture," *Pattern Recognition Letters*, vol. 11 (12), pp. 787–796, 1990.
- [3] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Journal of Optical Engineering*, vol. 32 (11), pp. 2824–2836, 1993.
- [4] A.P. Pentland, "A new sense of depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, July 1987.
- [5] M. Subbarao and G. Surya, "Depth from defocus : A spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.
- [6] D.T. Pham and V. Aslantas, "Depth from defocus using a neural network," *Pattern Recognition*, vol. 32, pp. 715–727, 1999.
- [7] M. Subbarao, "Parallel depth recovery by changing camera parameters," in *Proc. Second International IEEE Conference on Computer Vision*, Tampa, Florida, USA, 1988, pp. 149–155.
- [8] A.P. Pentland, T. Darrell, M. Turk, and W. Huang, "A simple real-time range camera," in *CVPR'89 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, June 4–8, 1989)*, Washington, DC., June 1989, pp. 256–261, Computer Society Press.
- [9] O. Monga, "Segmentation d'images : où en sommes-nous ?," Tech. Rep., INRIA, 1990.
- [10] J-P. Cocquerez and S. Philipp, *Analyse d'images : Filtrage et Segmentation*, 1995.
- [11] C. Simon, F. Bicking, and T. Simon, "Estimation of depth on thick edges from sharp and blurred images.," in *IEEE Instrumentation and Measurement Technology Conference*, Anchorage, USA, May 21-23 2002, IEEE, pp. 323–328.
- [12] H.H. Hopkins, "The frequency response of a defocused optical system," in *Proc. Royal Soc.*, 1955, vol. 231, pp. 91–103.